



MIT International Center for Air Transportation

Trajectory Clustering and Classification for Characterization of Air Traffic Flows

Mayara Condé Rocha Murça
Prof. John Hansman

November 5, 2015

Introduction

- **Increasing availability of massive air transport system data**
 - Radar tracks
 - Weather
 - Flight plans
 - Schedule
- **Needs for improved system capabilities aiming to increase efficiency of air traffic operations**
 - Post-event analytics
 - Monitoring and alerting
 - Real time decision support





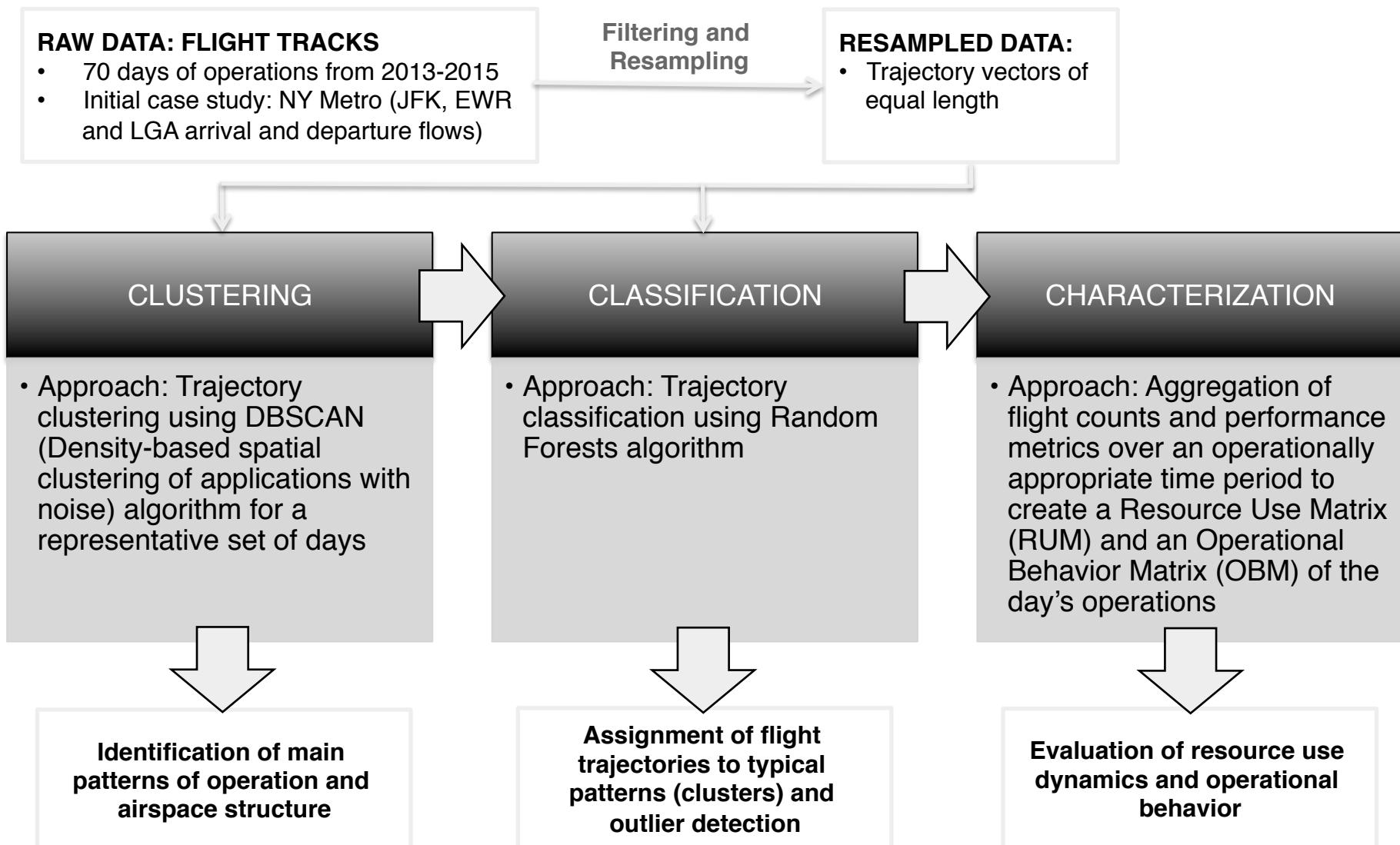
Research Objectives

- **Developing a data mining framework for characterization of air traffic flows:**
 - Identification of airspace structure
 - Assessment of conformance of flight trajectories against typical patterns
 - Evaluation of resource use dynamics and operational behavior under various operating conditions (demand, weather, TMI,...)

Input: ETMS track data

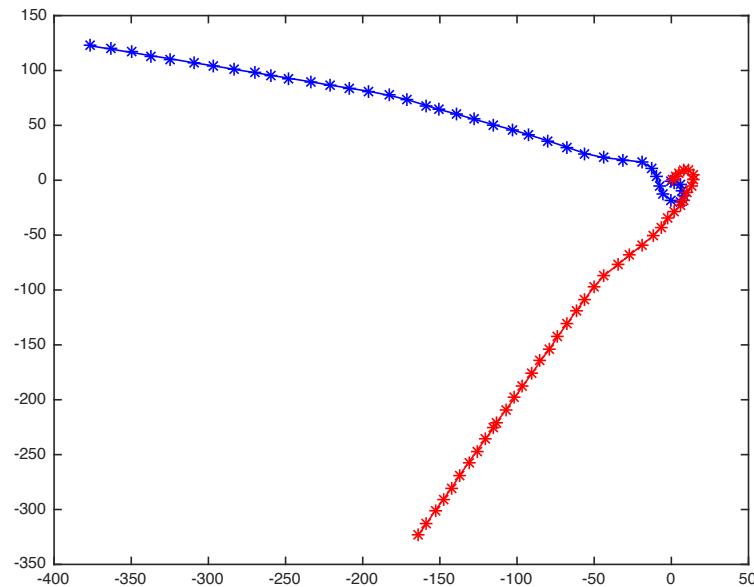
Flight index	Flight id	Time	Latitude	Longitude	Altitude	Speed	Origin	Destination	Aircraft
337780	VRD001	1436896800	38.04	93.84	36900	473	SFO	DCA	A320
344776	DAL2075	1436896800	38.10	90.66	31000	507	MCI	ATL	MD90
345401	SKW804T	1436896800	40.72	88.46	27400	429	ORD	SAT	E170
342237	AAL2069	1436896860	34.87	80.54	9100	279	JFK	CLT	A320
345010	SWA42	1436896860	32.25	86.92	40000	466	ATL	MSY	B737
345179	RPA4491	1436896860	35.77	84.18	29000	479	BNA	CLT	E170

Methodology



Trajectory Filtering and Resampling

- Flight times up to one hour
- Consideration of spatial and temporal dimensions
- Each trajectory is represented by a fixed length vector of features
 - Linear interpolation of 2D spatial position for a fixed number of points in time equally spaced along the duration of the flight



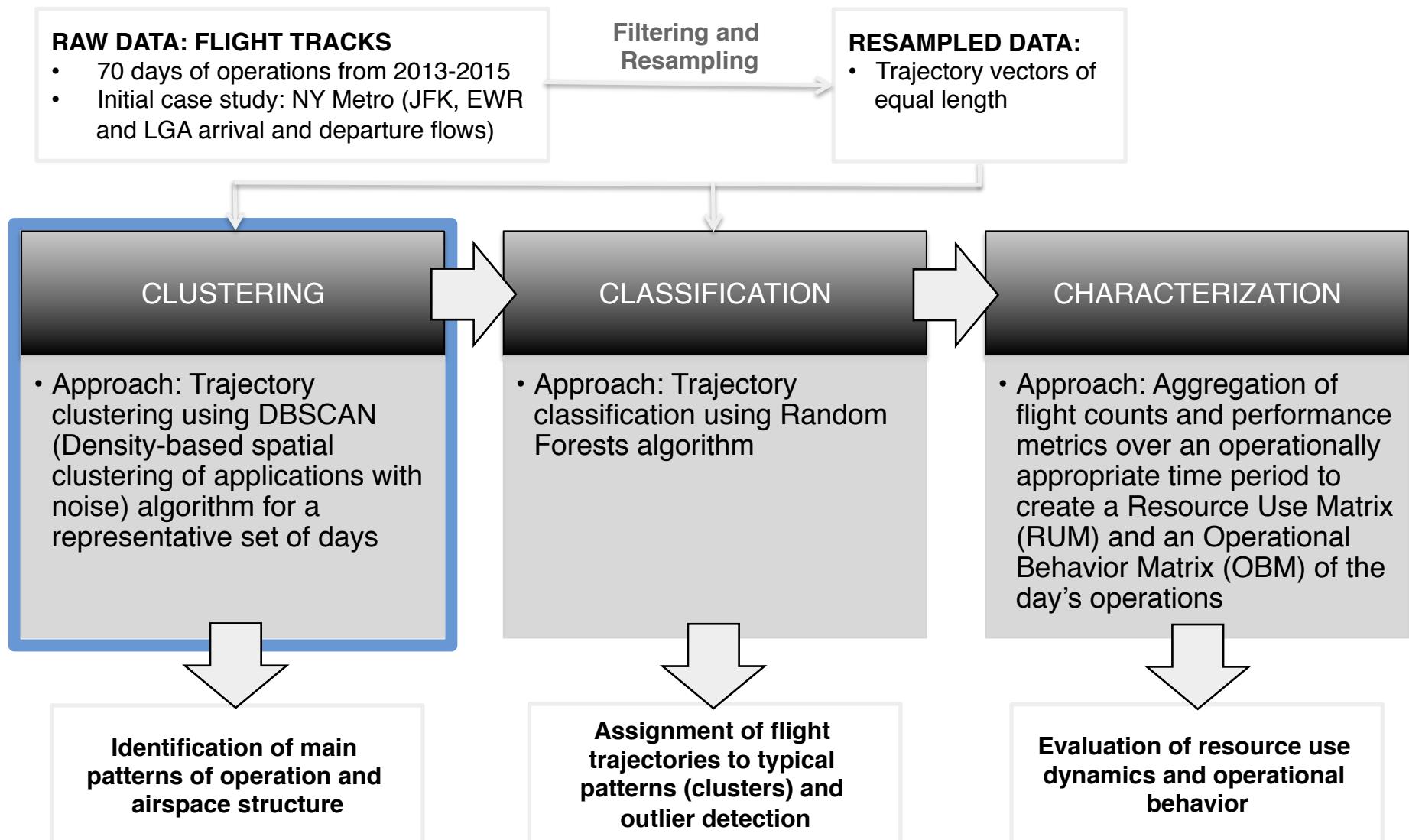
$$t_1 = (x_{11}, y_{11}, x_{12}, y_{12}, \dots, x_{1n}, y_{1n})$$

$$t_2 = (x_{21}, y_{21}, x_{22}, y_{22}, \dots, x_{2n}, y_{2n})$$

⋮

$$t_m = (x_{m1}, y_{m1}, x_{m2}, y_{m2}, \dots, x_{mn}, y_{mn})$$

Methodology



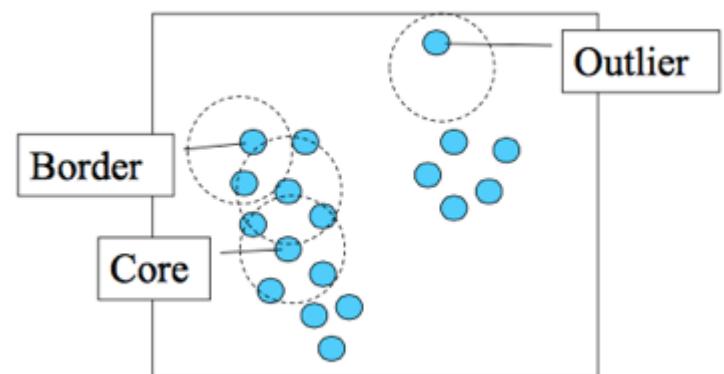
Clustering Analysis: DBSCAN

- **Basic idea:** a cluster is determined by a set of density-connected points in the data space
- **Advantages:**
 - Outlier detection
 - Discovery of non-convex clusters
 - No need to select the number of clusters a priori
- **Two input parameters:**
 - *Epsilon* (*Epsilon*-neighborhood):
 - ◆ $N_{Eps}(p) = \{q \in D / dist(p, q) \leq \varepsilon\}$
 - *MinPts*

Core point: It contains more than *MinPts* in its *Epsilon*-neighborhood

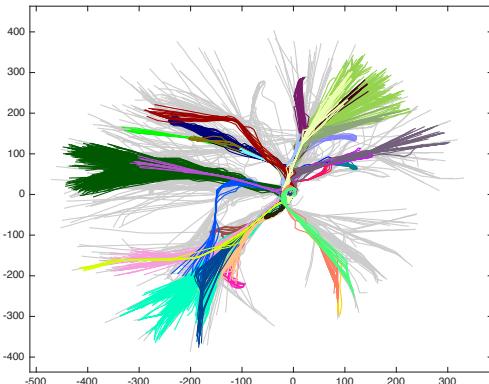
Border point: It is density-reachable from a core point

Noise point (outlier): It is not density-reachable from any other point in the database

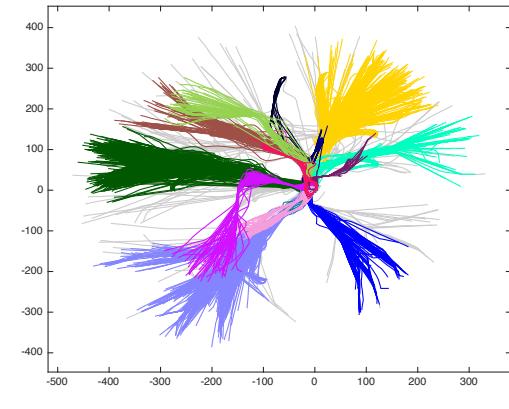
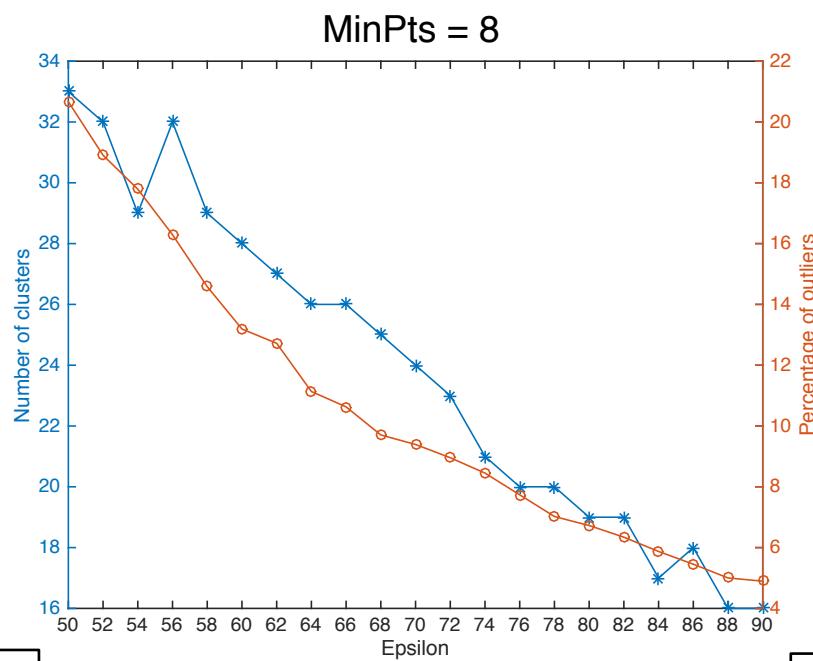


Clustering Analysis: DBSCAN

- Parameter selection was based on:
 - Sensitivity analysis and cluster validity indices (Davies-Bouldin Index and Silhouette Index) evaluation

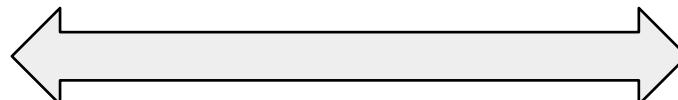


- Epsilon = 50
- 33 clusters
- 21% outliers



- Epsilon = 90
- 16 clusters
- 5% outliers

• Higher number of outliers
 • Undesirable cluster built from minor variations

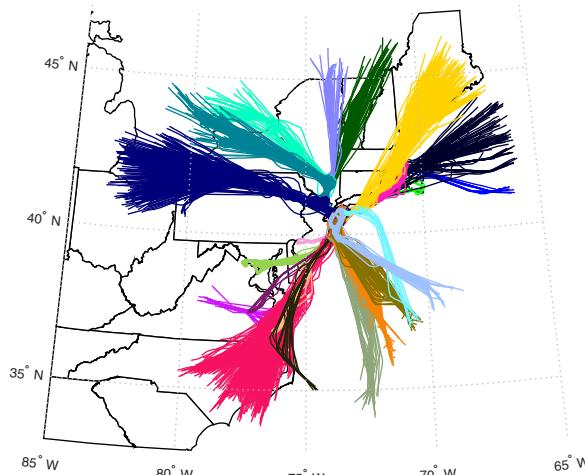


• Smaller number of outliers
 • Undesirable merger of clusters

Identification of Airspace Structure

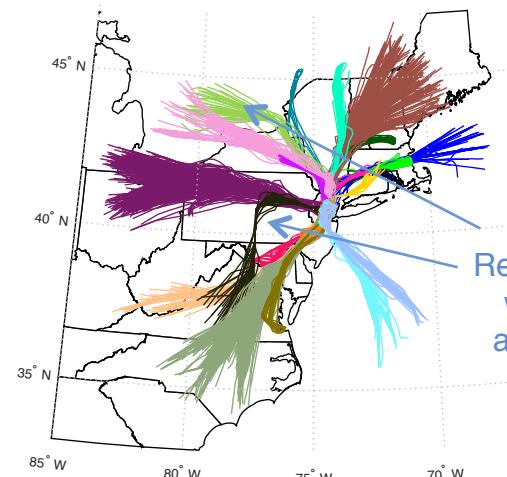
- Clustering results for arrival trajectories based on a dataset of 13 days (8 fair weather days and 5 convective weather days from 2013-2015)**

JFK Arrivals



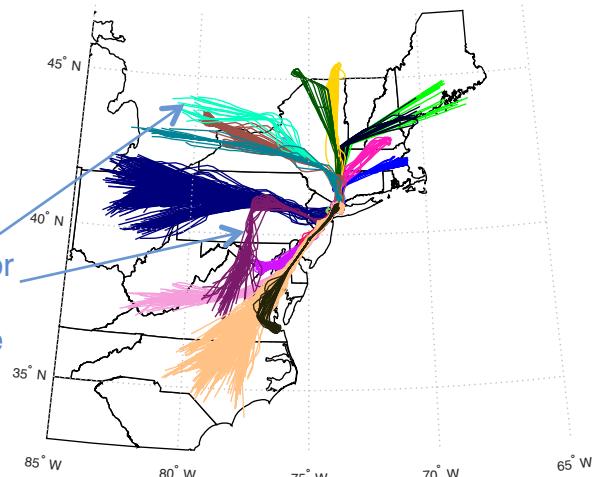
MinPts = 8
Epsilon = 73

EWR Arrivals



MinPts = 8
Epsilon = 72

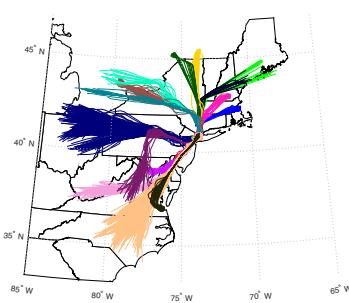
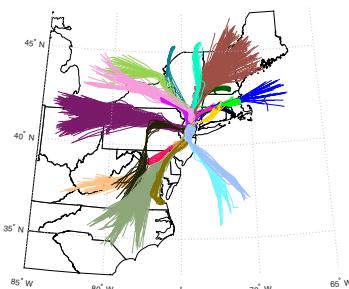
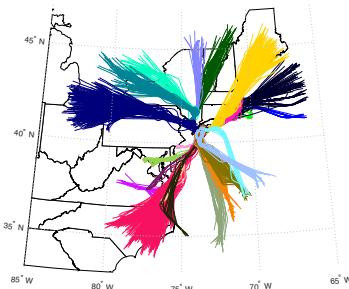
LGA Arrivals



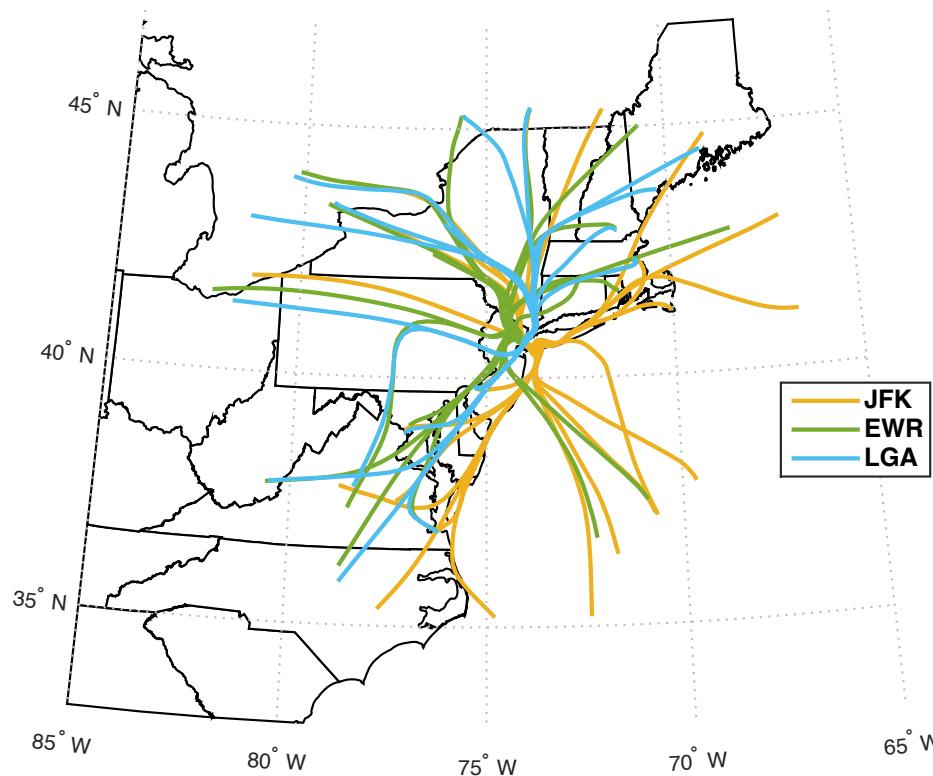
MinPts = 8
Epsilon = 68

Identification of Airspace Structure

- Clustering results for arrival trajectories based on a dataset of 13 days (8 fair weather days and 5 convective weather days from 2013-2015)



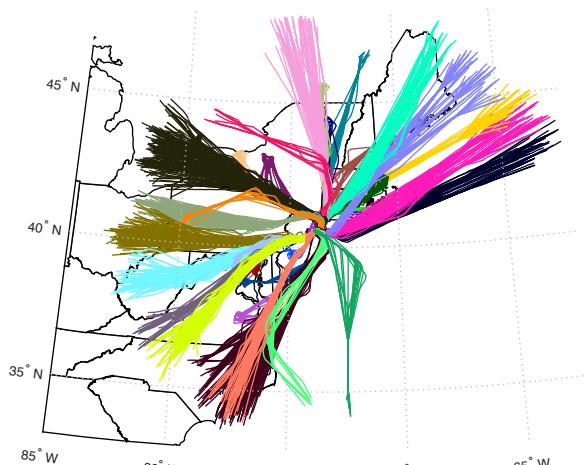
CENTROIDS OF NY ARRIVAL FLOWS



Identification of Airspace Structure

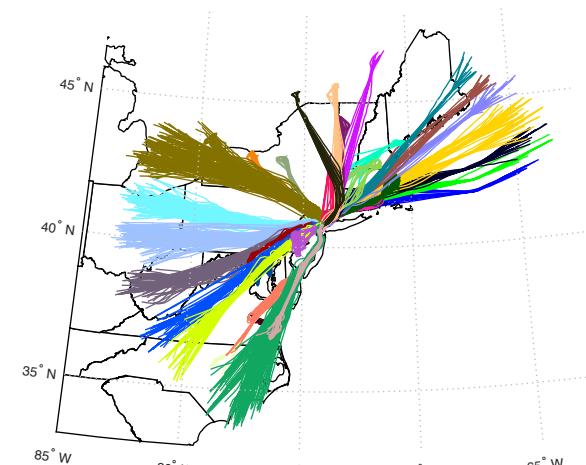
- Clustering results for departure trajectories based on a dataset of 13 days (8 fair weather days and 5 convective weather days from 2013-2015)

JFK Departures



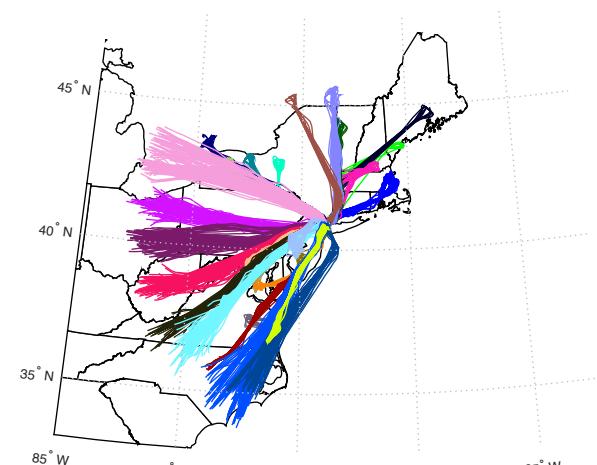
MinPts = 8
Epsilon = 68

EWR Departures



MinPts = 8
Epsilon = 63

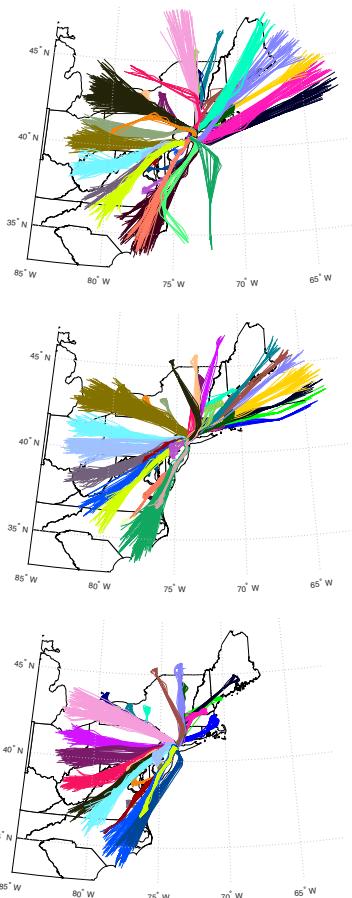
LGA Departures



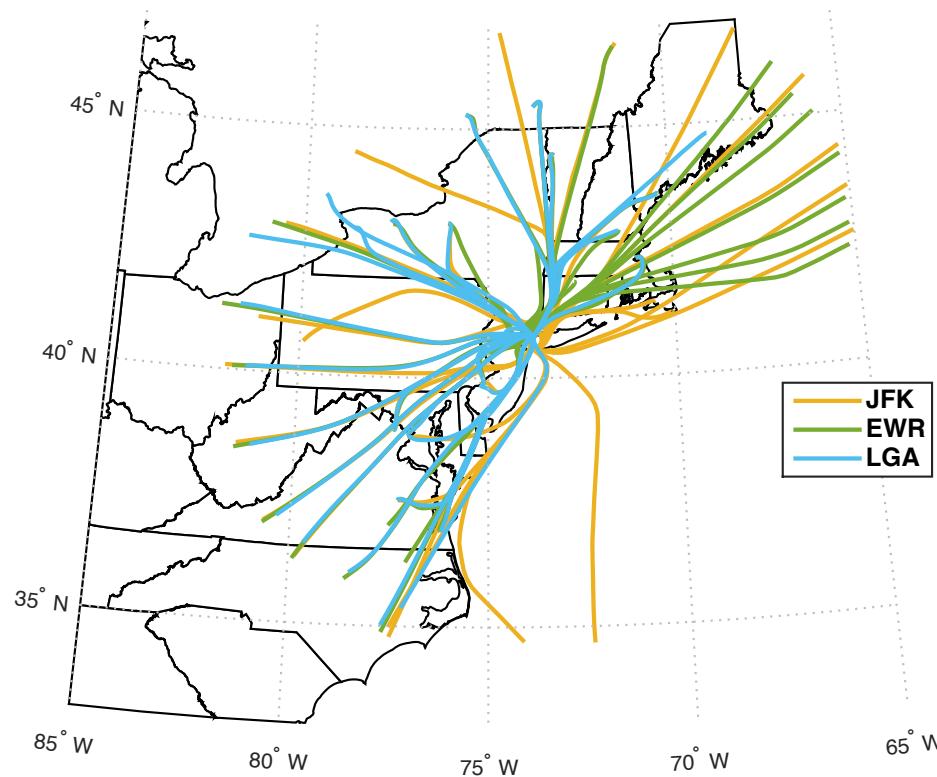
MinPts = 8
Epsilon = 54

Identification of Airspace Structure

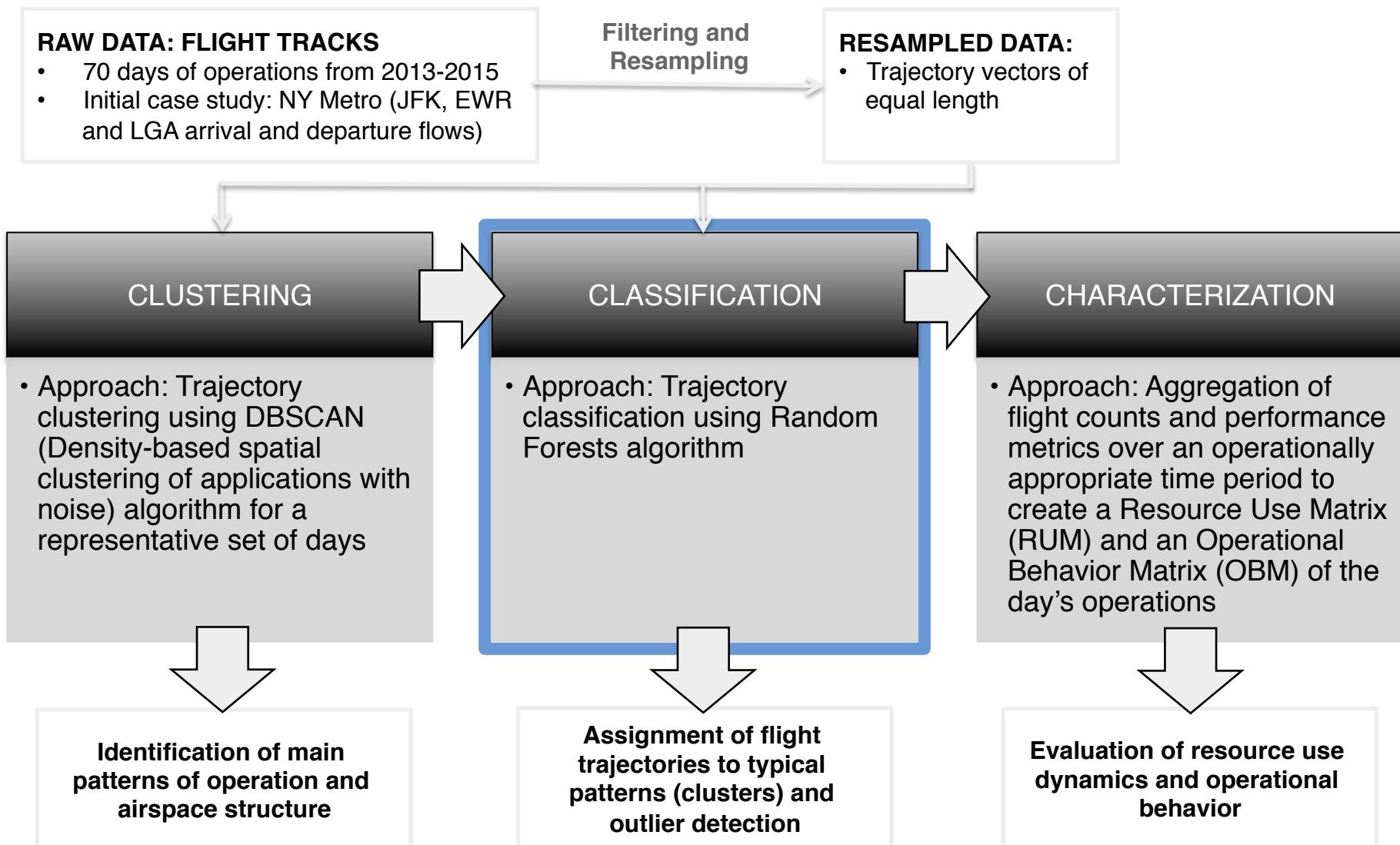
- Clustering results for departure trajectories based on a dataset of 13 days (8 fair weather days and 5 convective weather days from 2013-2015)



CENTROIDS OF NY DEPARTURE FLOWS

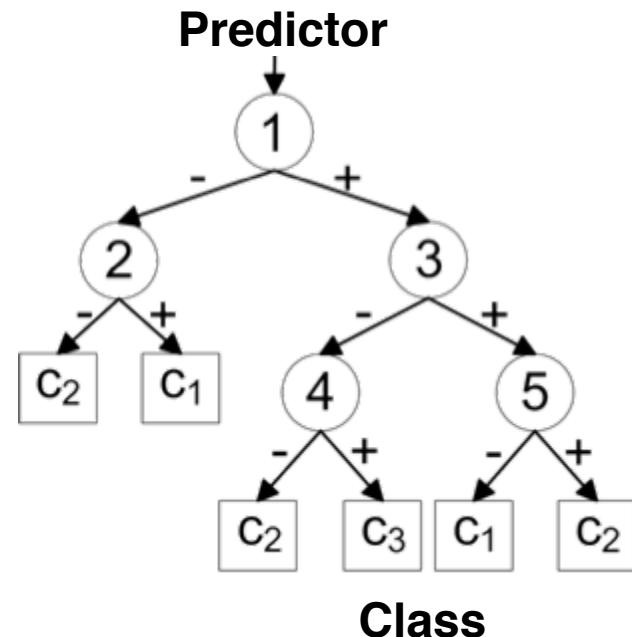


Methodology



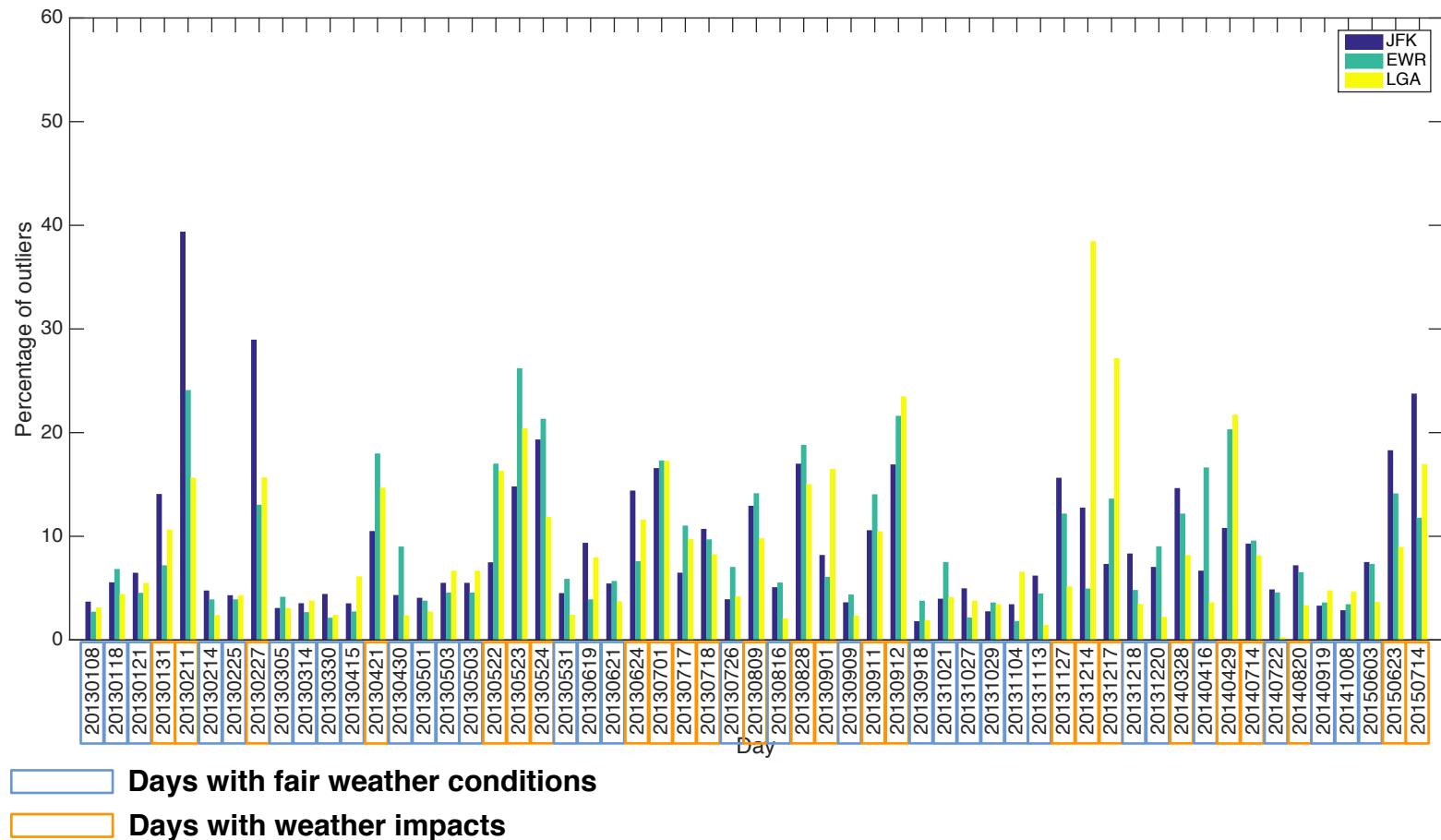
Classification Scheme: Random Forests

- **Basic idea:** a class is determined by the majority of votes from an ensemble of decision trees created from bootstrap samples of the data
- **Advantages:**
 - Fast computational time
 - High accuracy
 - Few parameters to tune
 - ◆ Number of trees
 - ◆ Node size
 - ◆ Number of predictors sampled



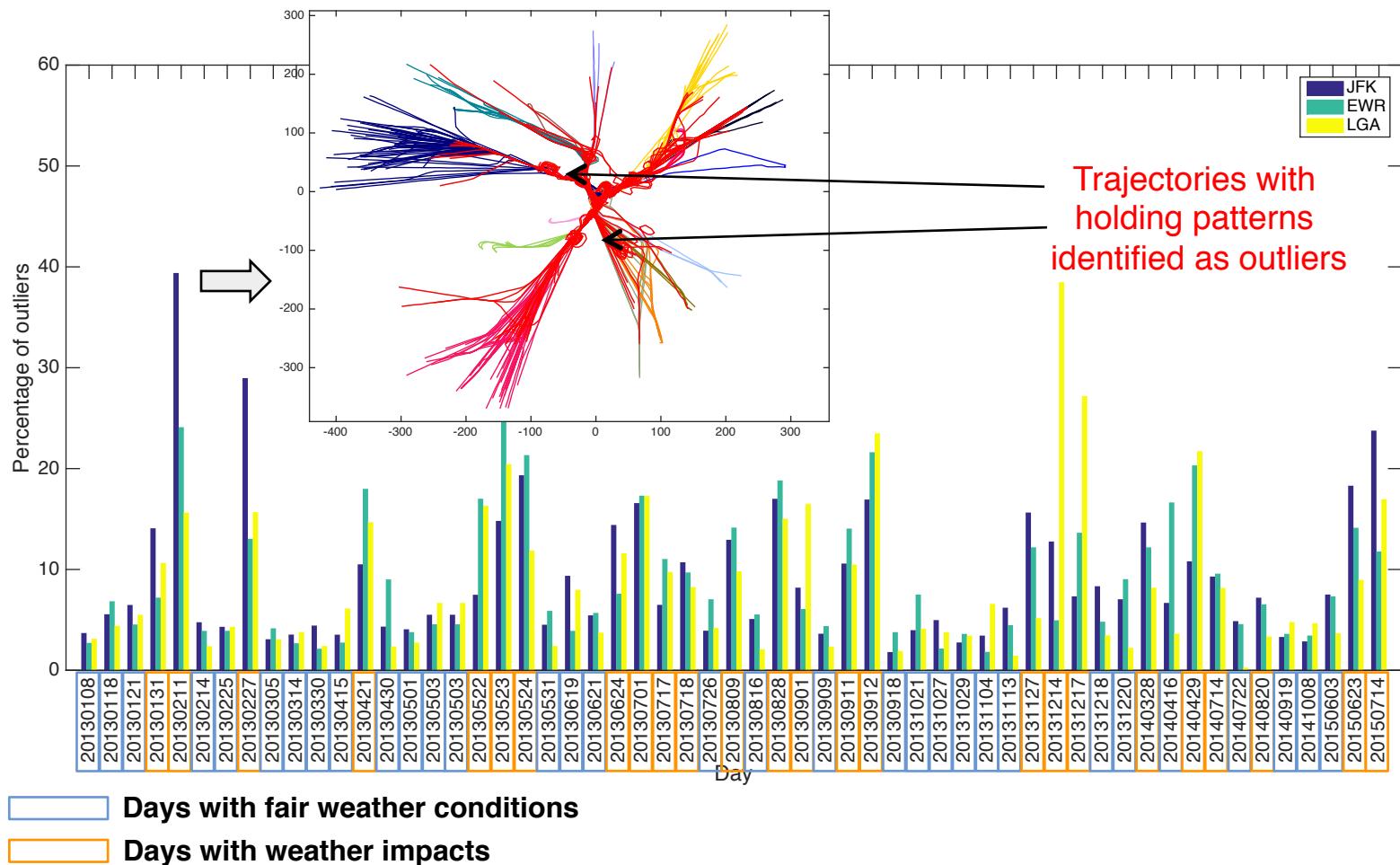
Flight Trajectory Conformance

- Classification results for arrival trajectories for 57 days (varied weather, demand, TMI,...)



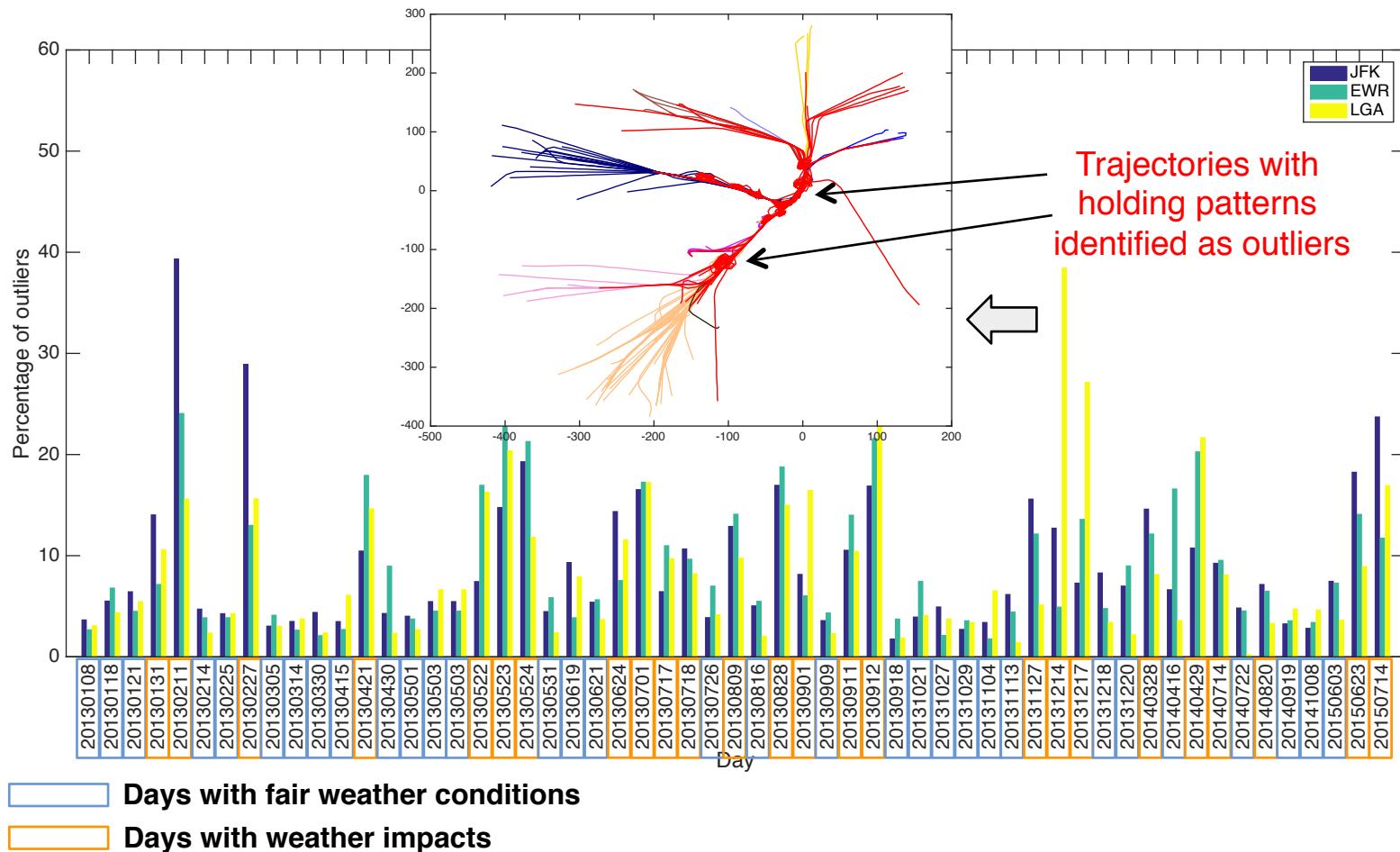
Flight Trajectory Conformance

- Classification results for arrival trajectories for 57 days (varied weather, demand, TMI,...)

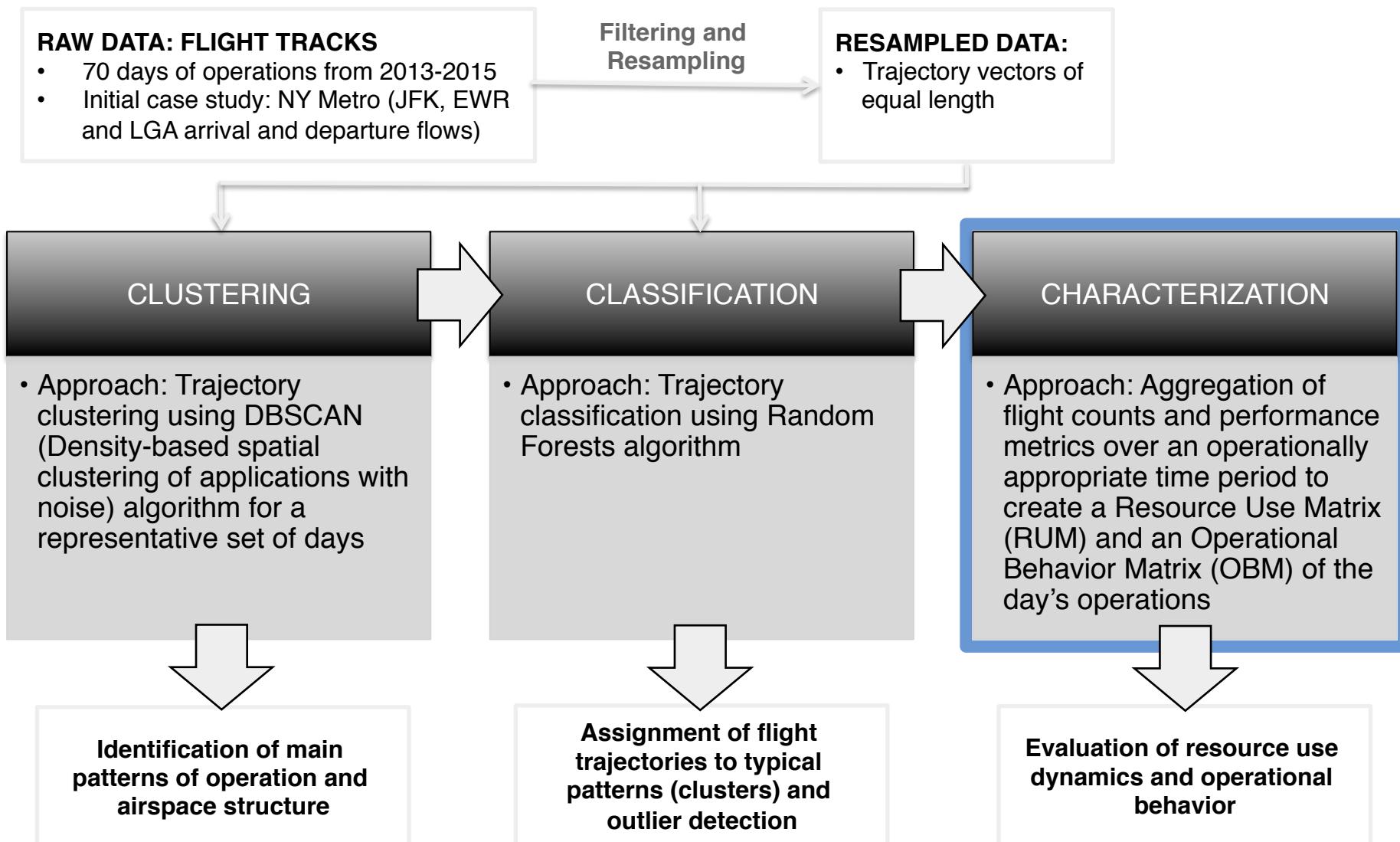


Flight Trajectory Conformance

- Classification results for arrival trajectories for 57 days (varied weather, demand, TMI,...)

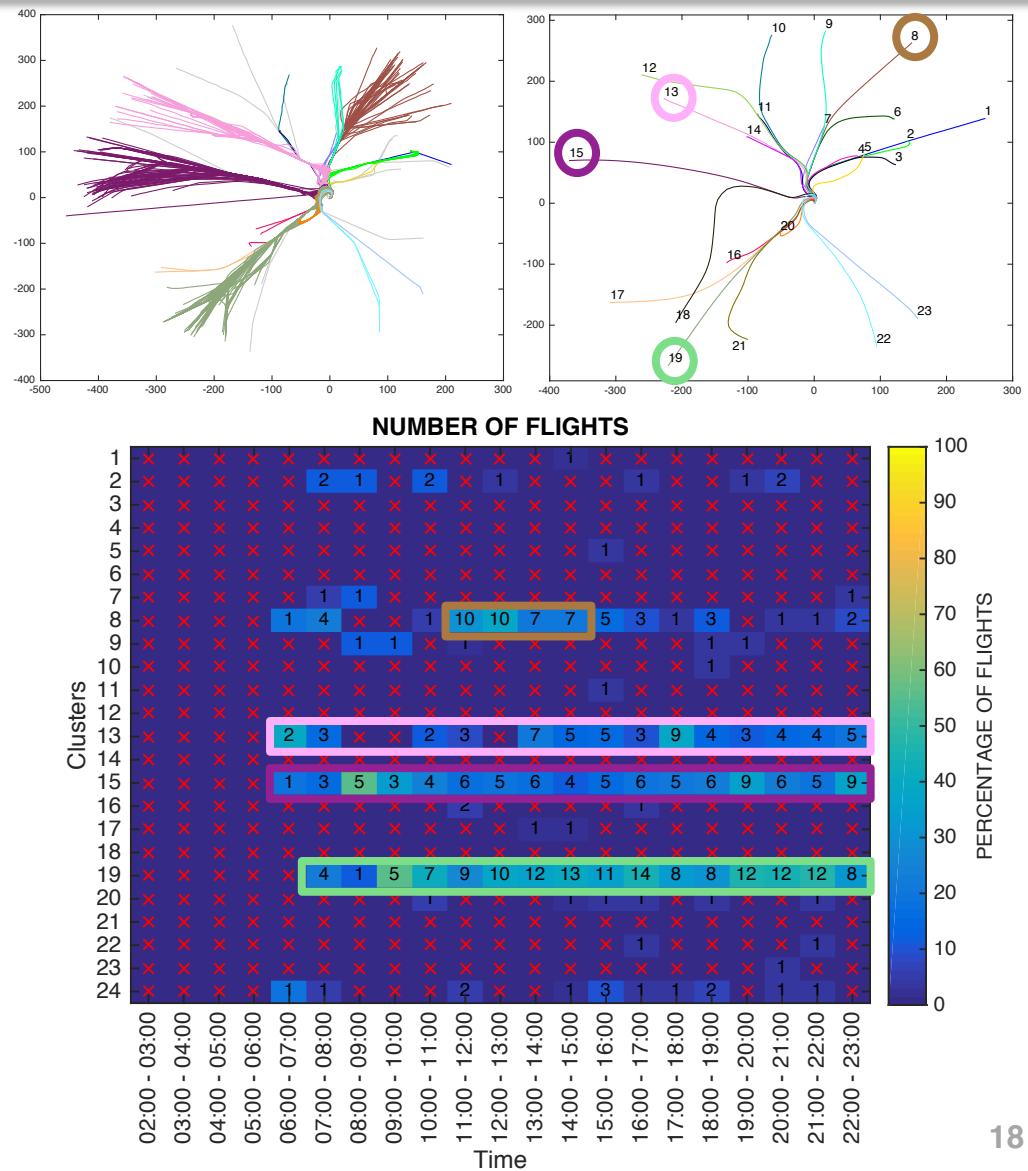
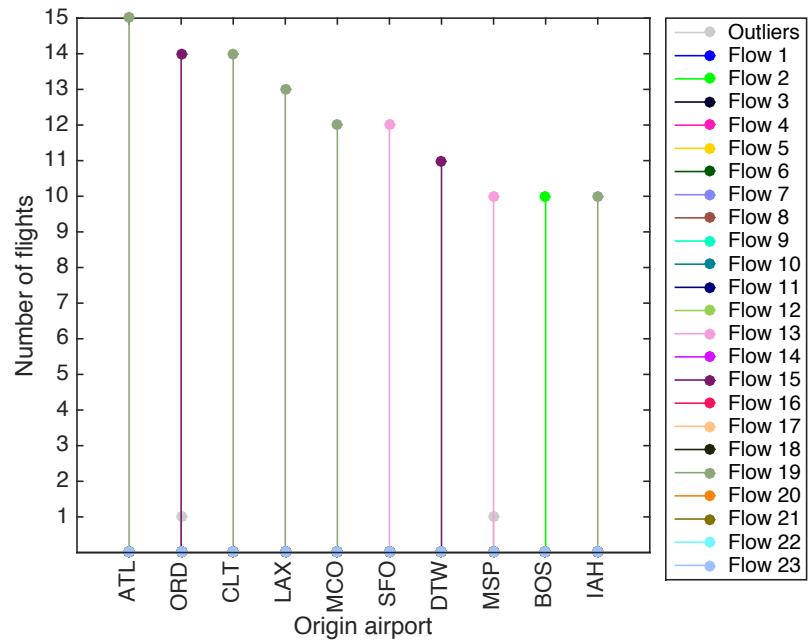


Methodology



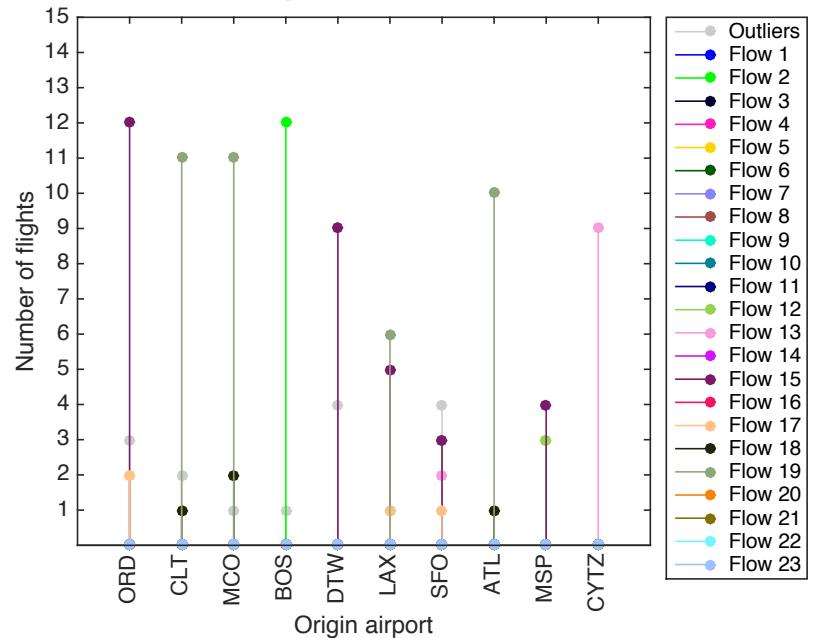
Resource Use Dynamics

- Example: EWR arrival flows for a clear weather day (October 8, 2014)**



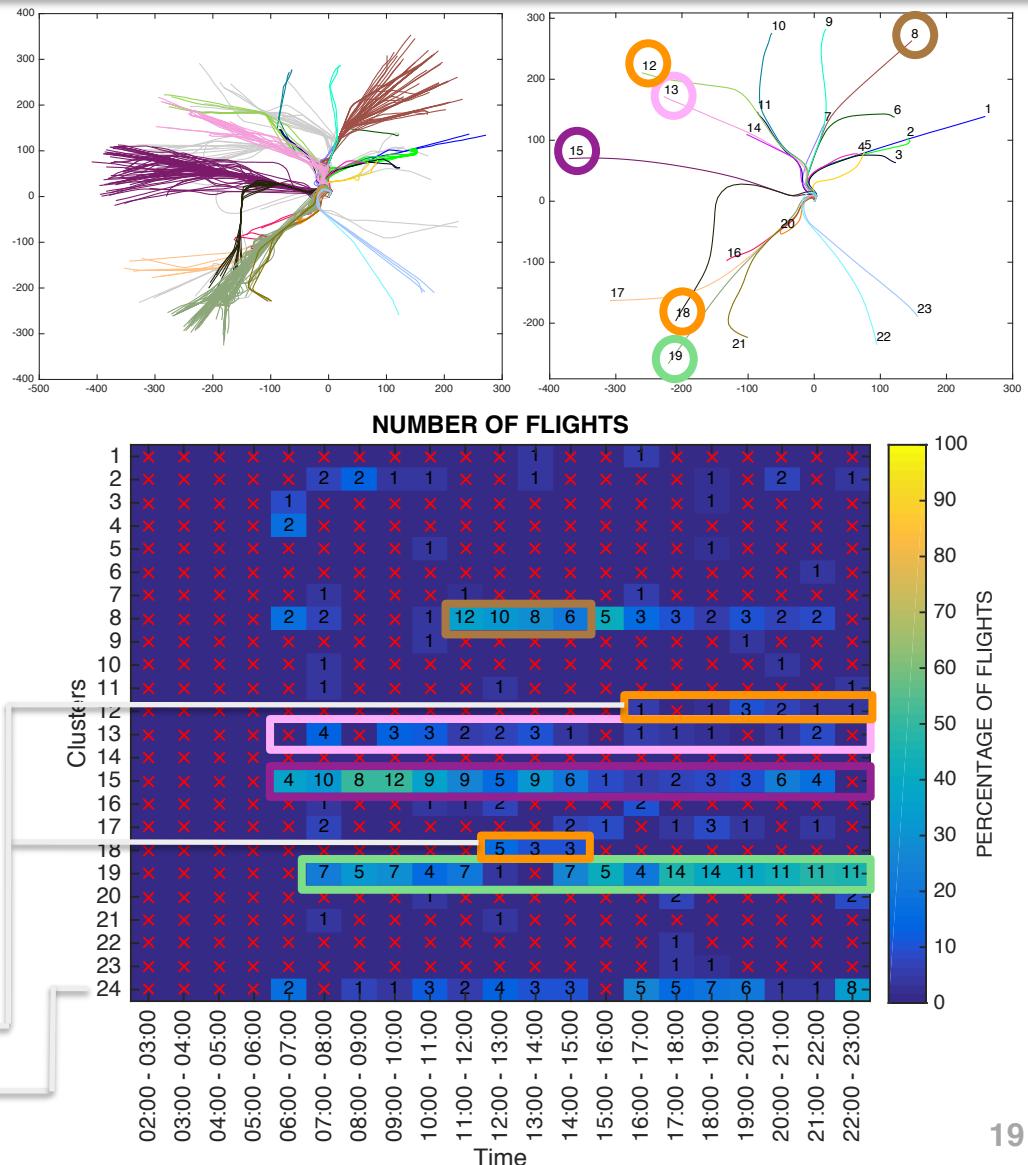
Resource Use Dynamics

- Example: EWR arrival flows for a convective weather day (July 14, 2015)**



Reroutes in effect

Assessment of outliers



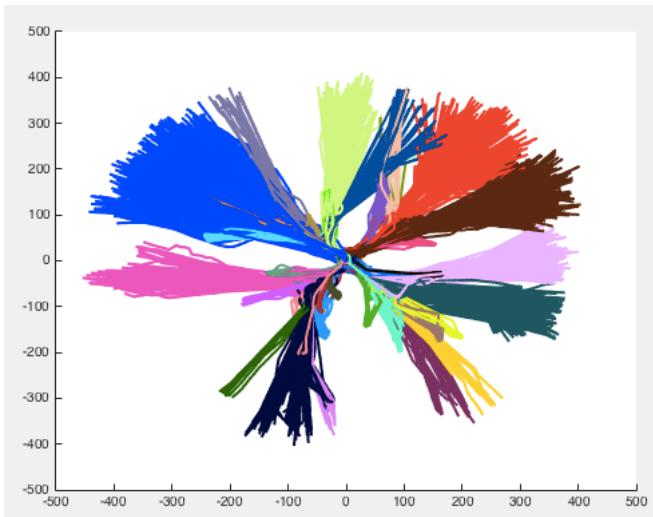


Conclusions

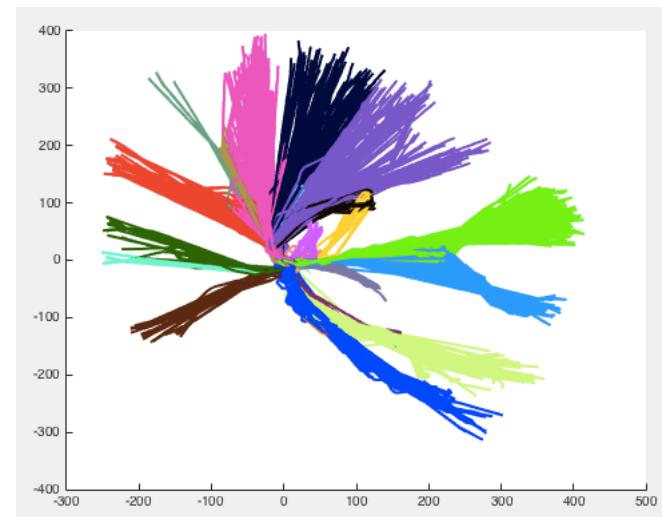
- **Development of a data mining framework for post-event characterization of air traffic flows:**
 - Identification of airspace structure from clustering analysis
 - Analysis of daily operations based on a classification scheme
- **Application to an initial case study (NY Metro)**
 - Identified typical arrival and departure flows to/from NY airports and shared resource use
 - Assessed flight trajectory conformance with respect to typical patterns and detected outlier trajectories
 - Identified temporal shifts in resource use

Next Steps

- Analysis of operational pattern recurrence for larger datasets and how they are correlated with demand, constraints (weather impacts) and performance (delays)
- Analysis of other airports and metroplex airspace



DFW Arrivals



SFO Arrivals



Thank you!
Questions?

mayara@mit.edu